# Grand Challenge Project

(http://www-rnc.lbl.gov/GC/)

D. Olson

RHIC Off-line Computing Review

30 July 1997

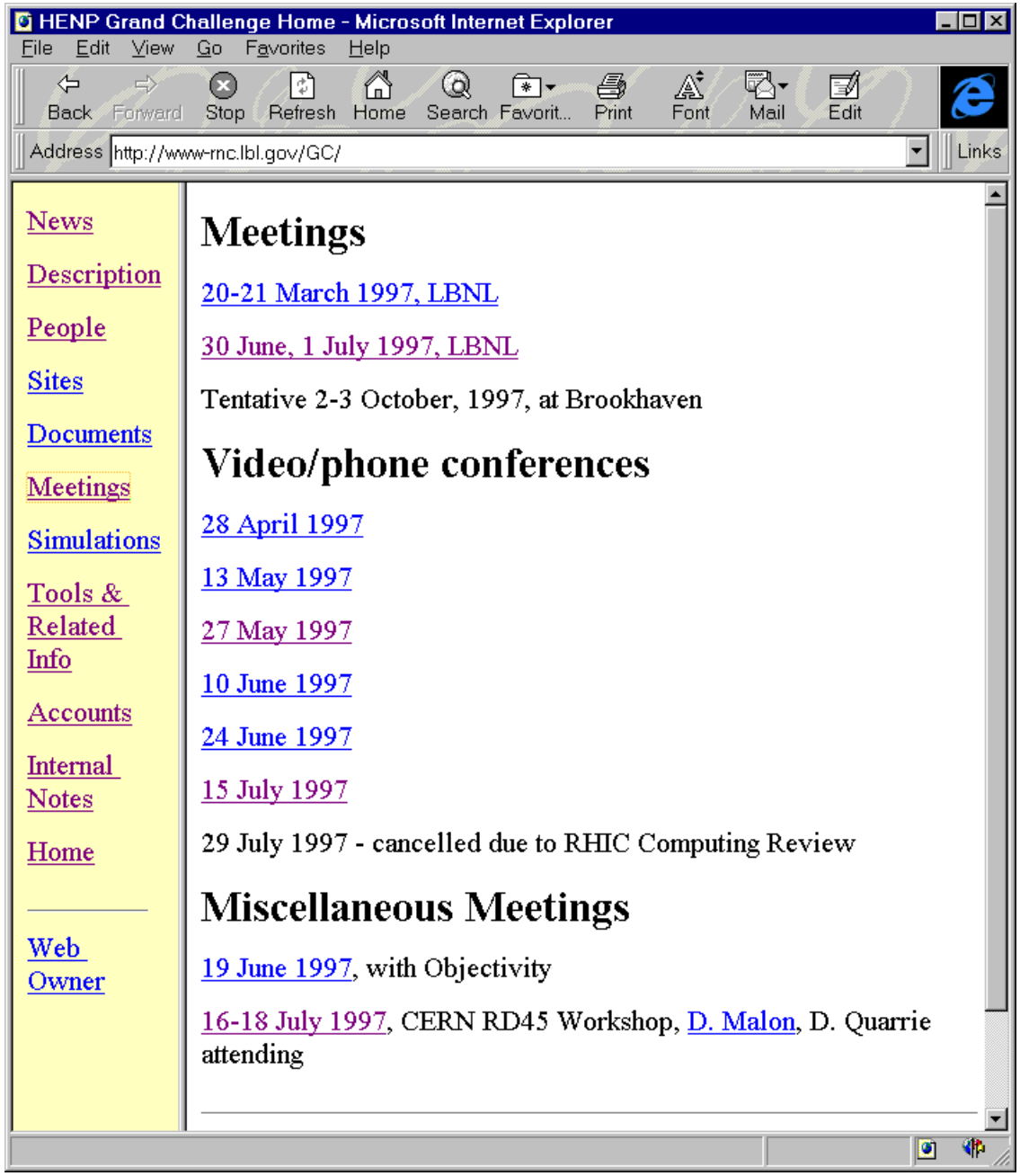
# Outline

- People
- Goals
- The Problem
- Approach
- Near Term Issues
- Schedule

# People (currently active)

| | | |
|---|---|---|
| LBNL | | |
| | NP | D. Olson (PI), G. Odyniec, F. Wang, N. Xu, R. Porter |
| | HEP | J. Siegrist (PI), I. Hinchliffe, R. Jacobsen |
| | Computing | C. Tull, D. Quarrie, W. Johnston,<br>A. Shoshani, D. Rotem, H. Nordberg |
| BNL | | |
| | RCF | B. Gibbard, D. Stampf, J. Flanigan |
| | Physics | D. Morrison |
| ANL | | E. May, D. Malon |
| FSU | | G. Riccardi |
| Rice | | P. Yepes |
| U.Tenn. | | S. Sorensen |

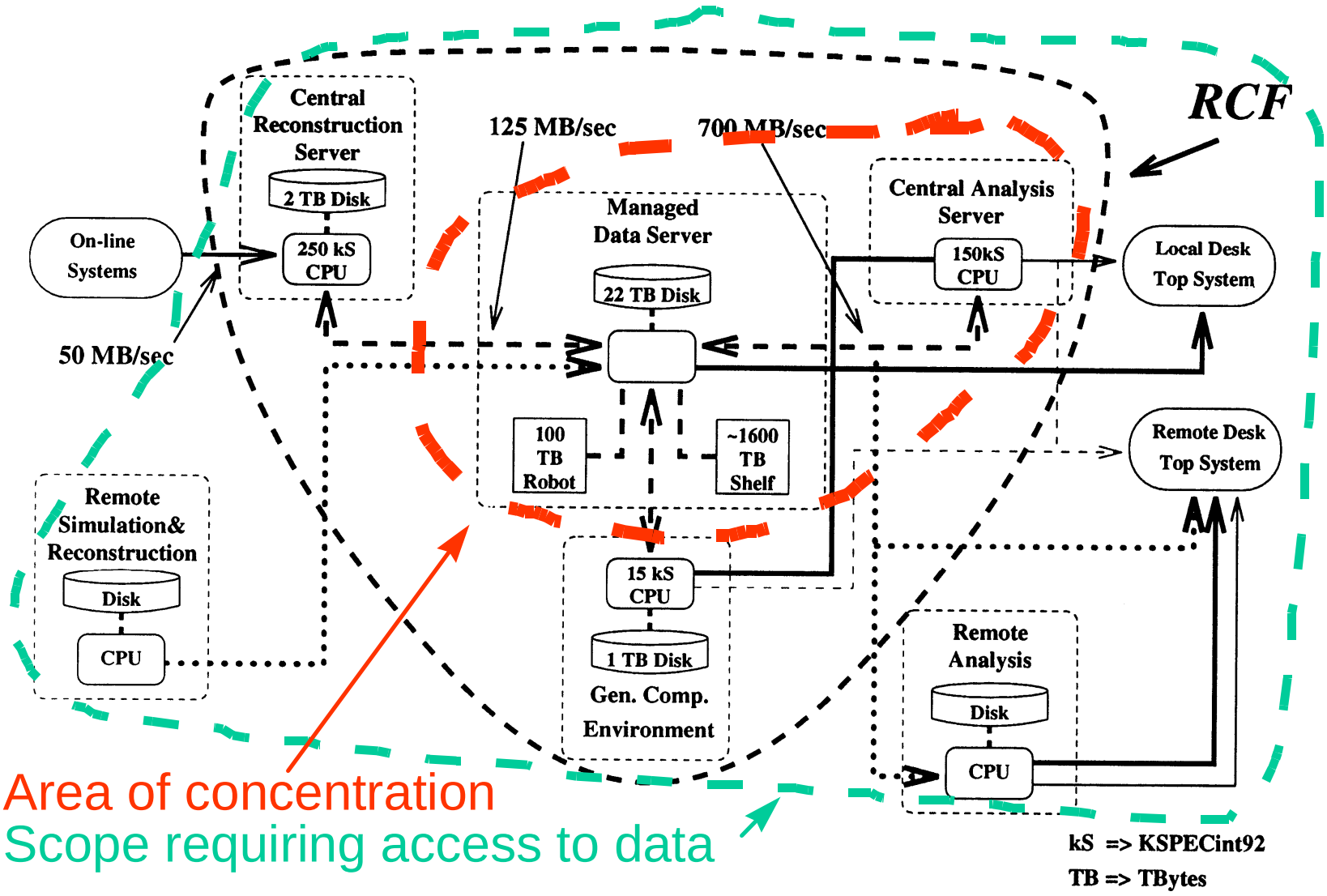
Expts: STAR, PHENIX, CLAS, BABAR, ATLAS


HENP Grand Challenge Home - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorit... Print Font Mail Edit
Address http://www-rnc.lbl.gov/GC/
Links
News
Description
People
Sites
Documents
Meetings
Simulations
Tools & Related Info
Accounts
Internal Notes
Home
Web Owner
Meetings
20-21 March 1997, LBNL
30 June, 1 July 1997, LBNL
Tentative 2-3 October, 1997, at Brookhaven
Video/phone conferences
28 April 1997
13 May 1997
27 May 1997
10 June 1997
24 June 1997
15 July 1997
29 July 1997 - cancelled due to RHIC Computing Review
Miscellaneous Meetings
19 June 1997, with Objectivity
16-18 July 1997, CERN RD45 Workshop, D. Malon, D. Quarrie attending

# Goals

- Demonstrate a solution for data access and analysis for RHIC.
- Three (2.5) year project (FY97, FY98, FY99).

# RHIC Computing Model



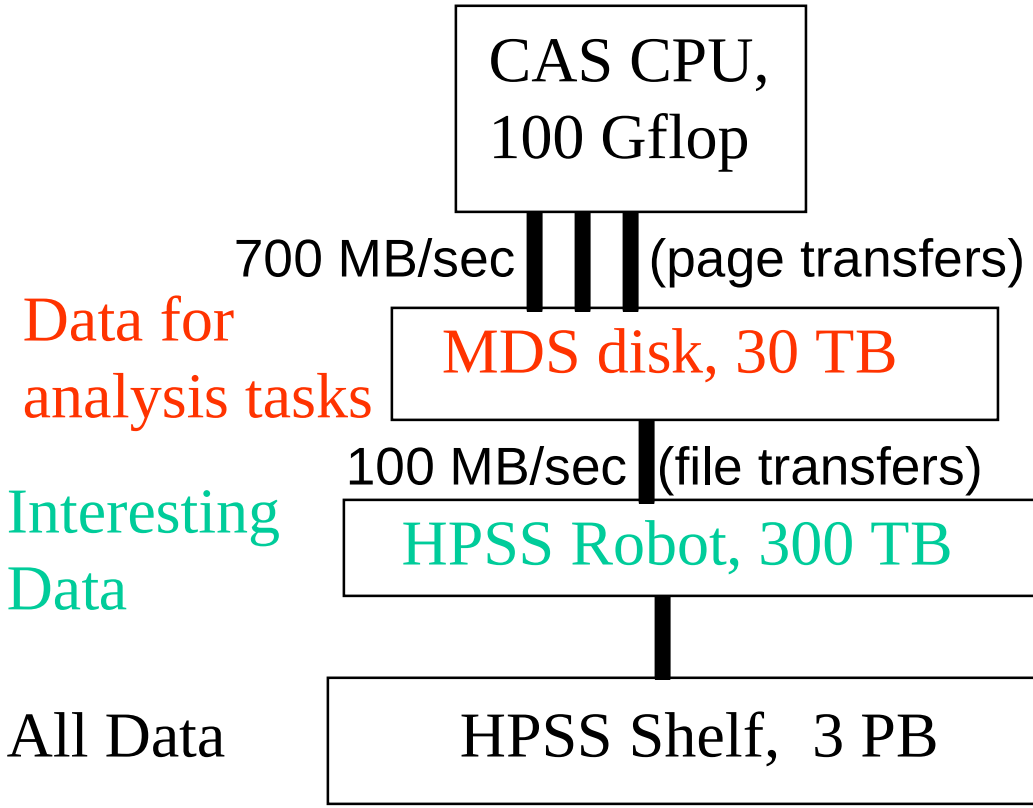
# Requirements

- Address the tape-disk-cpu data access bottlenecks.
- Achieve high-performance while maintaining human-efficient access to all data.
- Data access solution must not preclude requirements spanning RHIC computing:
  - event reconstruction (DST production)
  - selections (micro-DST generation)
  - analysis (single process development and PIAF-like parallel processing)
  - simulations (mixing data sources for comparison with theory)
  - robustness (operational efficiency > ??%)
  - tunable system (load balancing for op. efficiency)

# The Bottlenecks

(my est. for RHIC capacity, year 3, for scale)

Bulk bandwidth numbers meet estimated requirements assuming 100% efficiency.




How to achieve bulk bandwidth?

What fraction of data transfered is useful to programs?!!!

# Data organization & scheduling

- Define how to order files on tape.
- Define how to map substructures of events onto files (cluster by type).
- Define how order event (substructures) by feature, i.e., trigger streams, filtering, query patterns (cluster by value).
- Coordinate analysis tasks wanting data with the data available on disk.

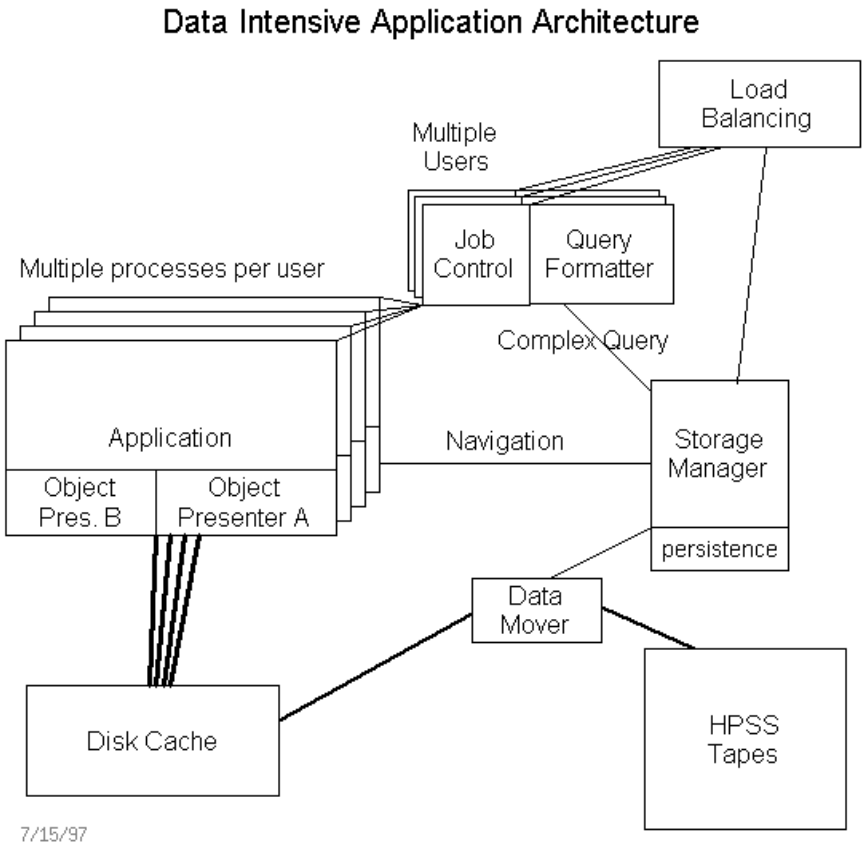
# Monitoring

- Items to monitor
  - File placement on tape.
  - Fraction of file accessed from disk.
  - Fraction of page used by program.
  - Bulk bandwidth used.
- Analysis of monitoring data is used to diagnose inefficiencies.
- System should be tunable based on this analysis.

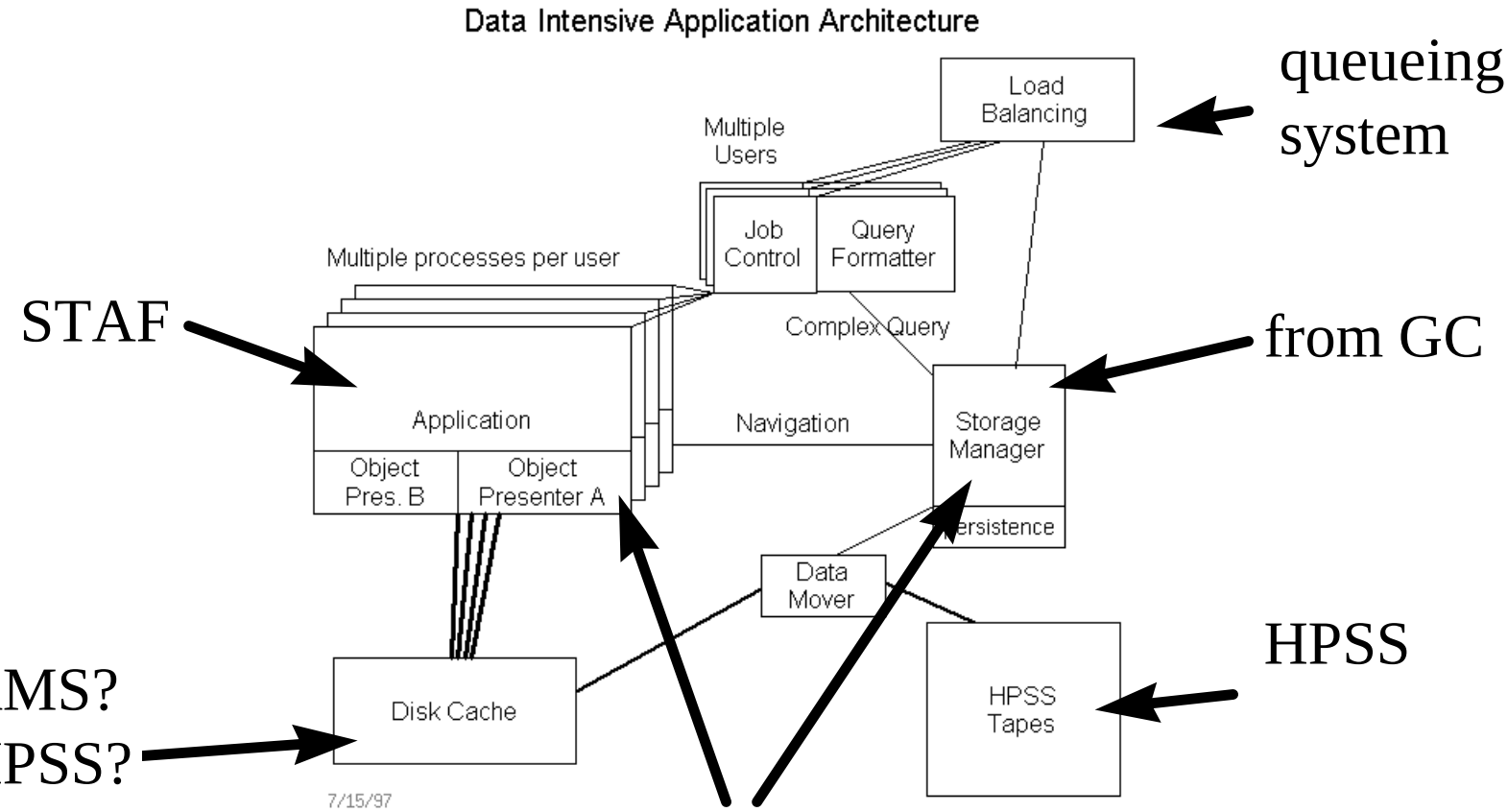
# The Approach

- Adopt an architecture which can address the year 2+ requirements.
- Develop early implementation which can meet year 1- requirements.
- Prototype at NERSC.
- Demonstrate at RCF some possible scenarios with simulated data.

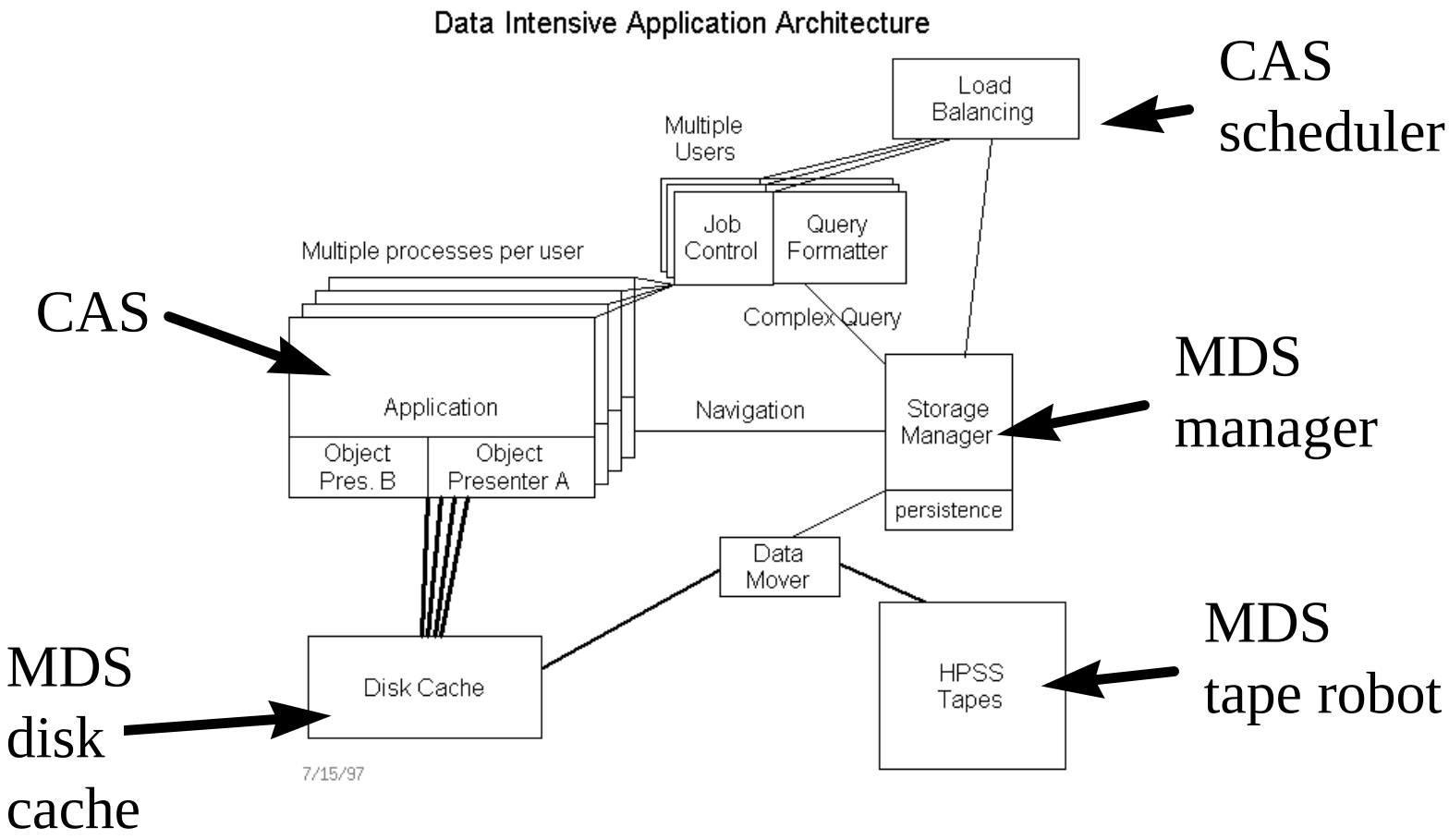
# The Architecture



# The Architecture (Software)



# The Architecture (Hardware)



# What's new in the data access approach

- ODMG model API for application code
  - like BaBar, RD45: a common HEP approach
  - great benefit to iterative development by maintaining object relationships across full dataset (majority of physicist time)
- Query, pre-fetch and query optimization
  - An object location index separate from the tape store enabling:
    - query-by-feature before touching tapes
    - ordering / scheduling access to files on tape
    - ordering / scheduling access to disk-resident objects
- Monitoring access efficiency
  - enables performance tuning via re-structuring and scheduling
- Data organization tools
  - enables re-structuring data for optimum access, where necessary

# Additional features of architecture

- Parallel event processing
  - PIAF-like event analysis
- Analysis framework (STAF) permitting mixed FORTRAN, C, C++ application code
  - with implications on the application level object model

# Issues: Software

- Objectivity/DB - role, scope, feasibility
  - evaluation
    - estimate time scale of feasible implementation
    - expect that distributed Objectivity federated DB unlikely in the near term
  - light-weight ODMG object presenter from ANL as alternative or additive until Objectivity is feasible?

# Issues: Hardware Testbed at NERSC

- In process of defining requirements.
  - Should support s/w development.
  - Should support enough performance tests to answer implementation questions like:
    - Objectivity?
    - Cost of re-organizing data?
    - HPSS disk vs. external disk cache?
    - Analysis tasks as direct HPSS clients?
    - Effect on CAS architecture?

# Schedule

3/97 - 9/97 Define architecture & Requirements

6/97 - 12/97 Technical choices & tests

6/98 First complete implementation of architecture

6/98 - 9/98 Test first implementation

9/98 - 12/98 Revise implementation

1/99 - 3/99 Test second implementation

4/99 - 6/99 Revisions & fixes

7/99 - 9/99 Perform final performance benchmarks

# Near-term plans

- Develop dataset of simulated events
- Collect data organization ideas from experimental groups
  (define query/access patterns)
- Investigate HPSS <--> disk issues.
- Investigate ODMG & Objectivity issues.
- Interface STAF to Objectivity.
- Implement prototype of architecture.

# Initial Software Prototype